

# Deleting Content

## Overview

There are three ways to delete content from an index:

1. Deleting by Document ID.
2. Deleting by query.
3. Deleting the entire index.

The sections below elaborate on the details of these operations.

[View incoming links.](#) >>

[Configure the Hadoop Cluster](#) , [Connectors](#) , [Content Ingestion - Concepts and Tools](#) , [Frequently Asked Questions](#) , [Indexes](#) , [JDBC Database Scanner](#) , [Joining Database Scanner](#) , [Load Data and Content](#) , [Loading AVRO Content](#) , [Loading CSV Content](#) , [Loading Excel Content](#) , [Loading File Content](#) , [Loading Parquet Content](#) , [Multiple Docs per Large XML File](#) , [Multiple Docs per Small XML File](#) , [One Doc per XML File](#)

- [Overview](#)
- [Delete a Single Document by Document ID](#)
  - [Delete Document Connector \(Single\)](#)
- [Delete Multiple Documents by Document ID](#)
  - [Delete Document Connector \(Multiple\)](#)
  - [Delete Documents Connector Properties](#)
- [Delete Documents by Query](#)
  - [Delete by Query Connector](#)
  - [Delete by Query Connector Properties](#)
  - [Performance Impact of Delete-by-Query](#)
- [Deleting the Entire Index](#)
  - [Delete Index from Administration UI](#)
  - [Delete-By-Query Connector](#)
  - [Delete Data-Agent Files](#)
  - [Delete HDFS and HBase Data](#)
- [Deleting from a Zoned Index](#)

## Delete a Single Document by Document ID

You can target a specific document for deletion if you know its document ID.

### Delete Document Connector (Single)

This procedure deletes a single document using a Delete Document Connector.

1. On the **Connectors** page of the AIE Administrator, click the New link and create a **Delete Document Connector**.
2. Give the connector a name
3. Enter the document ID in the **Document ID** field.
4. Enter the name of the workflow that was used to ingest the document.
5. Save the connector.
6. Run the connector to delete the document.

Scanner	Notes	Scheduler	Field Mappings	Advanced
Connector Name:		DeleteDocumentByID		
Node Set:		local		
Document ID:		CITY-Norway-Svolvaer		
Input File:				
Document ID Prefix:				
Ingest Workflow:		ingest		

Save Save & Test Cancel

# Delete Multiple Documents by Document ID

To delete multiple documents, create a text file with one document ID on each line.

## Delete Document Connector (Multiple)

This procedure deletes a list of documents using a Delete Document Connector.

1. On the **Connectors** page of the AIE Administrator, click the New link and create a **Delete Document Connector**.
2. Give the connector a name.
3. Enter the path/name of the file that contains the IDs. It goes in the **Input file** field.
4. Enter the name of the workflow that was used to ingest the documents.
5. Save the connector.
6. Run the connector to delete the documents.

com.attivio.connector.DeleteDocumentScanner

Scanner Notes Scheduler Field Mappings Advanced

Connector Name: DeleteDocumentByID

Node Set: local

Document ID:

Input File: c:\documents\DeleteIDs.txt

Document ID Prefix:

Ingest Workflow: ingest

Save Save & Test Cancel

## Delete Documents Connector Properties

The Delete Document Connector is configured by setting properties on the editor.

Delete Documents Connector Editor	Remarks
Connector Name	The name of the connector as seen in the UI or in XML.
Node Set	The nodeset the connector should run on. Defaults to default-service-nodeset. The Editor can set this value only on new, unsaved connectors.
Document Root	Xpath to the node that encloses the document, such as /rss/channel/item/.
Document ID	ID of the document to deleted (if only one).
Input File	Path to the file containing the list of document IDs to delete (if multiple). One document ID per line.
Document ID Prefix	Append this prefix to the Document ID during processing.
Ingest Workflow	Ingestion workflow to receive the ingested documents. String.

The table above is for the Delete Documents Connector Scanner Tab. The other tabs in the Connector Editor are described on the [Connectors](#) page.

## Delete Documents by Query

The **Delete by Query Connector** lets you submit a query to AIE as a string, and then deletes all of the documents that were returned by the query.



The delete-by-query connector's feedback message says it added zero documents to the index. This is true; it did add zero documents. It is not possible for a connector to report the number of documents that have been matched (and then deleted) by a query.



There is no way to "undo" the deletions. Be sure to test your deletion query beforehand by running it as a regular query and checking the documents returned.

## Delete by Query Connector

This procedure lets you delete the documents that match a query.

1. On the **Connectors** page of the AIE Administrator, click the New link and create a **Delete by Query Connector**.
2. Give the connector a name.
3. Enter either a [simple](#) or [advanced](#) query in the **Query** field.
4. Set the **Query Language** to "simple" or "advanced" to match your query's syntax.

 Do not change the **Query Workflow** default without good reason. The delete query must first be processed using query-side linguistic features, after which it must be placed in an ingestion pathway in order to be processed. The default value is usually appropriate.

5. Optionally, set the **Ingest Workflow** to **indexer** to bypass most of the ingestion workflow stages and reduce overhead.
6. Save the connector.
7. Run the connector to delete the matching documents.

**DeleteByQueryScanner: New Connector**

Scanner
Notes
Scheduler
Field Mappings
Advanced
Other

\*Connector Name:

Namespace Id:

Node Set:   

Query:

Norway

Query Language:  

Query Workflow:

Zone Names:

Document ID Prefix:

Ingest Workflow:

## Delete by Query Connector Properties

The Delete by Query Connector is configured by setting properties on the editor.

Delete by Query Connector Editor	Remarks
Connector Name	The name of the connector as seen in the UI or in XML.
Node Set	The nodeset the connector should run on. Defaults to default-service-nodeset. The Editor can set this value only on new, unsaved connectors.
Query	Query to use for deleting documents.

Query Language	Simple or Advanced Query Language.
Query Workflow	Query workflow for processing query.
Document ID Prefix	Append this prefix to the Document ID during processing.
Ingest Workflow	Ingestion workflow to receive the ingested documents.

The table above is for the Delete by Query Connector Scanner Tab. The other tabs in the Connector Editor are described on the [Connectors](#) page.

## Performance Impact of Delete-by-Query

Using a query to delete documents has a performance impact on AIE. Delete-by-query should be used cautiously and infrequently.

Delete-by-query, using complex queries, can cause performance problems in several ways.

First, note that newly-ingested documents are normally held in memory until written to disk as an index [segment](#). The system normally spaces out writing new index segments to disk, in order to avoid a major performance impact all at one time. Issuing a delete-by query will, in some circumstances (involving [Real-Time Updates](#), [JOIN queries](#), field roll-up, faceting, etc.), force AIE to flush all these newly-ingested in-memory documents to disk across all partitions before the delete query can proceed. This disk activity halts ingestion and querying during the flush.

In addition, note that the index will now contain additional (often small) new segments. This situation will cause the following performance issues:

- An index with a larger-than-usual number of segments will suffer slowed query performance until [background merging](#) corrects the problem.
- Background merging itself also has a performance impact.
- These new segments, created either by flushing or by merging, are copied to replicated indexes. This additional copying causes extra network usage. The more replicates and partitions there are, the more network bandwidth is required.

For all of these reasons, it is not a good practice to initiate many delete-by-query events in rapid succession.



### Best Practice

We recommend batching your delete-by-query events by ORing their conditions into a single query. This minimizes the number of flushing, merging, and replication events while still deleting the same records from the index.

The impact of delete-by-query is not directly related to the number of documents in the system, but background merging and replication take longer in larger systems.

Note that using a very simple query as the basis for deleting documents may not show these performance problems.

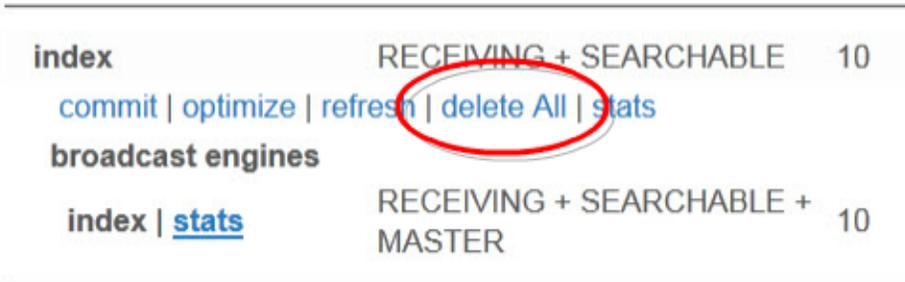
## Deleting the Entire Index

While developing new ingestion components, you will frequently need to delete all documents from the index and try again. There are multiple ways to do this:

- Use the **AIE Administrator** to delete all content from an index.
- Use a **Delete-By-Query Connector** to delete individual records that match a specific criterion.
- For single-node projects, you can delete the entire **data-agent** directory, which includes the index.
- For clustered projects, you can delete the index folders from HDFS and delete various utility tables from HBase.

## Delete Index from Administration UI

In the [Use the Attivio Administrator](#), navigate to **Monitors > Platform Components > Indexes**. Look for the **delete all** link in the index header.



When you use this link, AIE will ask you to type the full name of the index you want to empty.

## Delete-By-Query Connector

Using the AIE Administrator, it is quite easy to create a **Delete by Query Connector** using the wildcard query `*:*`. If you run this connector, it will delete all records in the index.

In practice, we advise against implementing this solution because it would be too easy for someone to run this connector accidentally.

## Delete Data-Agent Files

During the early stages of project development, when you are experimenting with alternate topologies and adjusting how to ingest data, there will be many times when it is convenient to remove all derived AIE data before running the system again. This avoids bugs caused by inconsistencies between the current experiment and the previous one.

For a single-node system, close the AIE-CLI, and stop the AIE Agent. Then delete everything in the **data-agent** folder. You will find the data-agent folder at `<install-dir>\bin\data-agent`, or at the location you specified when you started the [AIE Agent](#).

## Delete HDFS and HBase Data

If you are experimenting with a clustered (multi-node) project, you may sometimes need to clean out the data that AIE stores in HDFS and HBase. To remove data from previous runs, make an **ssh** connection to the Hadoop server (using PuTTY or Cygwin from Windows, if necessary). Then use these commands (note that if you changed the `hdfs.store.root` property, you should use the location you chose instead of `/attivio`):

### Linux server shell

```
sudo su hdfs
hadoop fs -ls /attivio
hadoop fs -rm -r -skipTrash /attivio/<prjName>
exit

sudo su hbase
hbase shell
list
disable_all 'attivio.*'
drop_all 'attivio.*'
exit
```

## Deleting from a Zoned Index

If you are using [Zoned Indexing](#), note that the **Delete Document Connector** (delete by document ID number) operates on the *default zone only*. It cannot reach documents in non-default zones.

The **Delete By Query Connector**, however, uses a query to locate the target documents. The query searches all zones by default, and can be adjusted to operate on specific zones if needed.